EFFECTS OF RESPONSE METHOD ON ERROR AND CYCLICAL BIAS IN PROPORTION JUDGMENTS

Andrew Morton and J. G. Hollands
Human-Computer Interaction Group
Defence Research and Development Canada – Toronto
Toronto, Ontario, Canada

The cyclical power model accounts for multi-cycle patterns of bias commonly observed in proportion judgments by proposing the use of intermediate reference points (Hollands & Dyre, 2000). We were interested in the effect of response method on the choice of reference points (fewer points lead to greater judgment error). Participants made estimates of proportions displayed in pie charts using one of three response methods: rotation of a dial, marker placement on a horizontal line, or a numerical estimate. Fitting the model indicated a two-cycle pattern for line and numeric conditions, but a four-cycle pattern for the dial, leading to reduced error. Response method did not affect the estimated value of the Stevens exponent (0.83 on average). Competing explanations of stimulus-response compatibility and response method are considered. Implications for the design of display and control systems are discussed.

INTRODUCTION

Proportion judgments require an observer to estimate the proportion one quantity is of another, larger quantity. For example, one could present a bar divided into two sections and ask an observer to judge the length of one section relative to the length of the entire bar. Responses can be collected using a numerical estimate (e.g. 25%) or by dividing another stimulus to correspond (e.g. dividing a line in two). People make proportion judgments in many situations: when examining statistical graphs, reading a speedometer or tachometer, judging the angle of orientation of an approaching ship, or checking dials showing temperature and pressure levels in a nuclear power plant. When doing so, they often respond in different ways—by recording a number, by adjusting a control, or by reproducing the position of a display indicator on paper.

Although we assume proportion judgments are accurate, biases (over- and underestimation) are commonly observed. These biases have important implications for many real-world tasks. The patterns of these biases have been studied extensively (Hollands & Dyre, 2000; Varey, Mellers, & Birnbaum, 1990), resulting in the development of explanatory theories.

The results typically obtained in proportion estimation experiments show a cyclical bias pattern. For example, Varey et al. (1990) showed observers squares containing black and white dots and asked them to estimate the proportion of dots of one type. Their data showed overestimation of proportions less than 50% and underestimation of proportions over 50%.

Sometimes the bias pattern repeats. For example Spence (Spence, 1990; Spence & Krizel, 1994) found two cycle patterns (over-under, over-under) when observers made proportion judgements with conventional graphical forms like the pie chart. To account for such patterns, Hollands and Dyre (2000) developed the cyclical power model (CPM), which proposes that an observer compares the part to intermediate *reference points* in the display rather than always making judgements relative to the whole. By increasing the number of reference points available to the participant the frequency of the bias pattern also increases. More importantly, increasing the number of reference points tends to decrease the magnitude of the error. The CPM also asserts that the phase of the bias pattern results from the stimulus continuum being judged, as represented by the Stevens exponent β (Stevens, 1975), with over-then-under occurring when $\beta < 1$. The further the exponent deviates from unity, the greater the amplitude of the cyclical bias. Estimated β values for judgments of proportions shown in pie charts are about 0.8.

Hollands and Dyre (1997) had observers make proportion judgements with pie charts but varied the frequency of available reference points (tick marks placed around the circumference of the pie, and/or on a horizontal response line). Bias pattern frequency increased with the number of tick marks from 2- to 4-cycle, and error decreased. However, when the

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **2003**	2. REPORT TYPE	3. DATES COVERED **00-00-2003 to 00-00-2003**
4. TITLE AND SUBTITLE **Effects of Response Method on Error and Cyclical Bias in Proporation Judgments**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Defence Research and Development Canada - Toronto,Human-Computer Interaction Group,1133 Sheppard Avenue West,Toronto, Ontario M3K 2C9,**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES **Proceedings of the Human Factors and Ergonomics Society's 47th Annual Meeting, October 13?17, 2003, Denver, Colorado**		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**5**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

number of tick marks was increased from 4 to 8 bias pattern frequency did not increase, implying that the additional tick marks were not being used as reference points.

In the Hollands and Dyre (1997) experiment, participants estimated proportions by dividing a horizontal line into two parts corresponding to the parts of the pie chart. It is a widely accepted human factors design guideline that dial displays are more compatible with rotary controls and linear displays with sliding controls (e.g., Ely, Thomson, & Orlansky, 1963; Wickens & Hollands, 2000). Fitts and Seeger (1953) found that such stimulus-response (S-R) compatibility reduced response time and error. The implication is that Hollands and Dyre might have found an increased frequency in cyclical bias (and accompanying decrease in error) as the number of tick marks was increased to 8 if they had used a circular response method.

Morton and Hollands (2002) examined this question by having participants respond using a rotary dial. Tick marks were placed at regular intervals on the pie and/or dial at the same frequencies used by Hollands and Dyre (1997). Generally, a four-cycle bias pattern was produced regardless of the number of available reference points, leading to the hypothesis that a response method has an inherent set of reference points and thus different frequencies would result from the use of different response methods. The purpose of the current experiment was to examine directly the effect of response method on cyclical bias patterns in proportion judgments. In particular, we predicted that greater compatibility between display and response method should increase the number of reference points used and reduce judgment error.

METHOD

Participants

Eighteen naive participants (mean age 28.3 yrs.) with normal or corrected-to-normal vision participated. An equal number of males and females were run in each condition. The participants were recruited from the staff of a scientific institute and students of local universities.

Apparatus and Stimuli

Stimuli were presented on a 43 cm (21") CRT monitor at 1280 x 1024 pixel resolution. A graphics workstation controlled the display and collected participants' responses. Pie charts of diameter 7.2 cm were presented at screen center. The proportion displayed in the graph was varied. Proportions representing 5 to 95% in divisions of 5% and multiples of eighths and sixteenths were used (31 different levels in total). One portion of each graph was shown in blue, the other in green.

There were three response method conditions. In the dial condition, participants used a rotary dial (diameter 5 cm, height 2.5 cm) mounted on the top surface of an input box. The dial had a radius line etched on its surface to indicate direction and a physical stop prevented the dial from being rotated through the 12 o'clock position. The input box contained an optical encoder whose output signal varied linearly with the position of the response dial. This signal was passed to the graphics workstation via a Cereal Box Model LV824-F-4e (BG Systems, Palo Alto, CA), which then computed a proportion. In the Line condition, a horizontal line (length 15.875 cm) was displayed below the pie chart. The participant clicked on the line using the mouse to place a response marker on the line. In the numeric condition, an input prompt was shown and the participant typed a percentage response using a numeric keypad. In all conditions the participant pressed the 'Enter' key to end the trial.

Design and Procedure

The experiment had a one-way between-subjects design with response method (dial, line, numeric) serving as the independent variable. 186 trials were formed by the factorial combination of 31 proportions x 6 repetitions. The 31 proportions were grouped together to form a block. The order of the proportions depicted within a block was randomized. There were 9 practice trials. Dependent measures were judged proportion and response time.

Participants were seated approximately 40 cm in front of the monitor. Each participant read the experimental protocol and gave informed consent. Participants read a set of written instructions on the monitor and completed practice and experimental trials. On each trial a pie chart divided into two sections was displayed. Participants in the dial condition were told to rotate the circular response dial so that the pointer line divided the circle into two parts whose relative sizes corresponded to the parts shown in the pie chart. Participants in the line condition were told to click on the line such

that the resulting mark divided the line into two parts whose relative sizes corresponded to the parts shown in the pie chart. Participants in the numeric condition were told to input a number from 0 to 100% to indicate the size of the green slice relative to the whole as shown in the pie chart. Each participant took about 20 minutes to complete the session.

RESULTS

Model Fits

A bias score (judged – true proportion) was computed for each trial. The CPM was fit to mean bias values for all participants in each condition using a nonlinear estimation procedure (see Hollands & Dyre, 2000, for details). Table 1 shows that the two-cycle model version produced the best fit (as determined by R^2 values) for the line and numeric conditions, but the four-cycle model version fit best for the dial condition. The best-fitting model versions are depicted in Figure 1.

The CPM was also fit to each participant's data individually (see Table 2). R^2-values from each participant's 2 and 4 cycle fits were submitted to a 3 x 2 (Response Method x Model Version) between/within analysis of variance (ANOVA), which indicated that the four-cycle model version produced a greater R^2 value than did the two-cycle for the dial condition, but that the reverse was true for the Line and Numeric conditions, $F(2,15) = 16.85$, $MSE = 0.029$, $p < .0005$.

Estimated Stevens exponents (β-values) from each participant's 2 and 4-cycle fits were submitted to a 3 x 2 ANOVA. The β-values for the better fitting model version (M = .80, .82, and .88, for Dial, Line, and Numeric, respectively) were generally less than those for the worse fitting model version, which were generally close to unity $F(2,15) = 4.32$, $MSE = 0.24$, $p < .05$. The values for the better-fitting model version did not differ from each other, Newman-Keuls, $p > .05$.

Absolute Error

An absolute error score (|judged proportion – true proportion|) was computed for each trial. A mean absolute error value was computed for each participant. Given that the CPM predicts reduced error with an increase in the number of bias cycles, and that higher frequency fits were found for dial than line and numeric response methods, mean absolute error data were submitted to two planned comparisons, each comparing the dial to one of the other two response methods. Error was smaller in the dial condition (M = 1.32) than in the line condition (M = 5.40), $F(1,10) = 26.99$, $MSE = 5.01$, $p < .0005$. Although mean error was greater for the numeric condition (M = 2.16) than the dial condition, the difference failed to reach conventional significance levels, $F(1,10) = 3.22$, $MSE = 0.666$, $p = .103$.

Response Time

A mean response time was computed for each participant in each condition. These data were submitted to a 3 x 2 ANOVA. No differences were observed in response time between conditions ($F < 1$, $p > .85$).

DISCUSSION

The best-fitting version of the CPM differed across response methods, suggesting that observers used different reference points with the different methods. In contrast, the best-fitting model version was identical regardless of whether data were averaged across participants or not. In addition, response method did not affect the estimated value of the Stevens exponent (0.83 on average). In the dial condition, where observers showed higher-frequency bias patterns, they were also more accurate than in the line or numeric condition (although the latter difference was not reliable).

We can consider two possible explanations as to why changes in response method affected cyclical bias. The first is based on S-R compatibility. The more compatible S-R relationship between pie chart and dial, relative to the other conditions, led to the use of higher-frequency reference points, thereby improving judgment accuracy. The second explanation is that reference points may be implicit in different response mechanisms, thereby influencing the frequency of response. A circular response dial, where a four-cycle pattern fit best, can be divided into four using Cartesian axes. However it is not clear why participants were reduced to fewer reference points with the line or numerical methods—why did these methods discourage the use of reference points at quarters, for example? Experimental work is underway to disambiguate these explanations.

This work has application in the design of display and control systems. While the CPM

potentially allows designers to optimize tick mark frequency, the results of this experiment highlight the importance of considering the display-control relationship, and demonstrate the utility of the CPM for detecting performance differences not evident with traditional error measures. Additional experiments are being conducted to investigate further the effects of different stimulus and response set combinations.

Table 1. Average Group Fits of the Cyclical Power Model (best-fitting version indicated).

Model Version	Response Device					
	Dial		Line		Numeric	
(# of Cycles)	R^2	β	R^2	β	R^2	β
1	-0.028	1.002	0.013	1.058	0.401	0.941
2	-0.028	0.995	0.392*	0.805*	0.506*	0.877*
4	0.817*	0.794*	-0.073	0.889	0.035	0.928
8	-0.030	1.004	-0.075	0.742	0.010	1.105

Table 2. Averaged Individual Fits of the Cyclical Power Model (best fitting version indicated).

Model Version	Response Device					
	Dial		Line		Numeric	
(# of Cycles)	R^2	β	R^2	β	R^2	β
1	0.022	1.002	0.121	1.073	0.126	0.942
2	-0.065	0.995	0.149*	0.821*	0.285*	0.879*
4	0.411*	0.797*	-0.005	0.896	0.003	0.930
8	-0.082	1.009	-0.077	0.760	-0.028	1.111

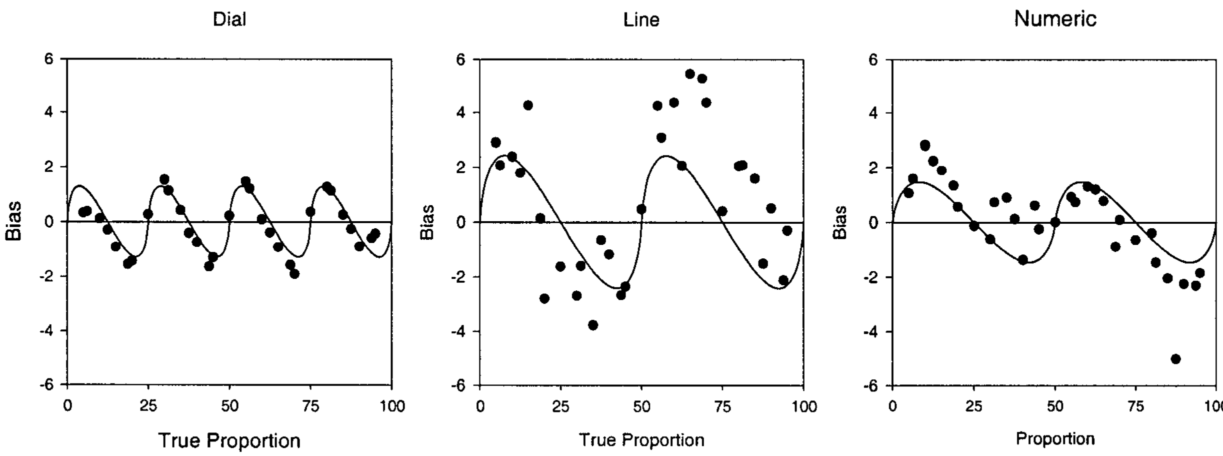

Figure 1. Bias results from each condition. Continuous functions represent the best fitting version of the CPM.

ACKNOWLEDGMENTS

This research was supported by a DRDC Technology Investment Fund grant to the second author. We thank Garry Dunn for programming support and John Bowen for construction of the circular response device. The first author is now at the Department of Mechanical and Industrial Engineering, University of Toronto.

REFERENCES

Ely, J. E., Thomson, R. M., & Orlansky, J. (1963). Design of controls. In C. T. Morgan, J. S. Cook, A. Chapanis, & M. W. Lund (Eds.), *Human engineering guide to equipment design* (pp. 247-280). New York: McGraw-Hill.

Fitts, P. M. & Seeger, C. M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology, 46*, 199-210.

Hollands, J. G. & Dyre, B. P. (1997). Bias in proportion judgements with pie charts: The cyclical power model. In *Proceedings of the Human Factors and Ergonomics Society-41st Annual Meeting* (pp. 1357-1361). Santa Monica, CA: Human Factors and Ergonomics Society.

Hollands, J. G., & Dyre, B. P. (2000). Bias in proportion judgments: The cyclical power model. *Psychological Review, 107,* 500-524.

Morton, A., & Hollands, J. G. (2002) [Stimulus-response compatibility in proportion judgments]. Unpublished raw data.

Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 683-692.

Spence, I. & Krizel, P. (1994). Children's perception of proportion in graphs. *Child Development,* 65, 1193-1213.

Stevens, S. S. (1975). *Psychophysics.* New York: Wiley.

Varey, C. A., Mellers, B. A., & Birnbaum, M. H. (1990). Judgements of proportion. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 613-625.

Wickens C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd Ed.). Upper Saddle River, NJ: Prentice-Hall.